Proceedings of the Institute of Acoustics

DIRECTIVITY OF SPHERICAL MICROPHONE ARRAYS

C.F. Cardoso Institute of Sound and Vibration Research, University of Southampton,
P.A. Nelson Highfield, Southampton SO17 1BJ, UK

1 INTRODUCTION

The aim of this research is to examine the performance of a microphone array mounted on a rigid sphere. When building the microphone array several features have to be considered, for example noise should not degrade the signal and no irrelevant talkers should be detected. Also, in large rooms reverberation is usually a problem and the microphone has to be able to select the main signal. For instance, in a teleconference the main signal is the person talking.

Numerical experiments were undertaken, within a MATLAB environment, by running computer simulations that tested the directivity of a rigid sphere microphone array in both near and far field. Scattering of sound by a rigid sphere was studied in order to understand its effects in the directivity of the array. The beamformer used in the study is a "focused beamformer" that enhances the microphone directivity. A focused beamformer attempts to map the distribution of acoustic source strength associated with a given source distribution by changing the beamforming operation in accordance with the assumed source position.

2 BACKGROUND

Interaction between parties at remote locations can be made more effective and efficient by using as much perceptual information as possible. In audio, this perceptual information is achieved by reproducing the voices of the talkers so that these are perceived as emanating from the positions associated with the visual stimuli[1]. When using audio-visual applications with several participants it is extremely difficult to relate the output of a single audio transmission to a specific talker.

A microphone array is one solution to this problem. A microphone array is a network of sensors whose output signals are combined in order to enhance the sound waves coming from a desired direction[2]. Microphone arrays have several applications, namely in speech data acquisition systems, such as teleconferencing, speech recognition and speaker identification, sound capture in adverse environments, and biomedical devices for the hearing impaired[4]. Audio is often used as a complementary modality to video data to provide faster localisation over a wider field of view[5-6]. By using an array of microphones a directional listening response can be achieved that rejects interfering noise from other directions. Early radio astronomy techniques demonstrated that a microphone array can be focused merely by the way in which the signals are processed with no mechanical adjustment[3]. Enabling hands-free acquisition of speech with little constraint on the user and not requiring physical movement to alter its direction of reception are also advantages.

Microphone array research has been undertaken specifically with regard to spherical microphone arrays capable of steering beams in multiple directions[7]. Using a three dimensional configuration eliminates forward-backward grating lobes which occur in two-dimensional arrays; also linear and two-dimensional arrays have limitations such as a restricted scan field and a narrow operational frequency bandwidth[8]. In addition, by being spherical full scan coverage is achieved, its aperture direction is independent making the beam-width and gain of the antenna constant for all scan angles[8]. Recent work demonstrated that the acoustic response of arrays where the microphones are mounted on a scattering body, for instance a rigid sphere, leads to a highly desirable array directivity[9-10-11].

3 MICROPHONE ARRAY

Figure 1 shows the block diagram that illustrates the simulations performed. A point monopole source radiates sound to the microphone array. When the signal arrives at the microphones it is scattered due to the presence of a rigid sphere. The microphone array then processes the resulting signal through an array of filters that ensure the time alignment of the signals at the summing junction. The output is then passed through an inverse filter with the goal of obtaining an output impulse response that is a delayed version of the source signal.

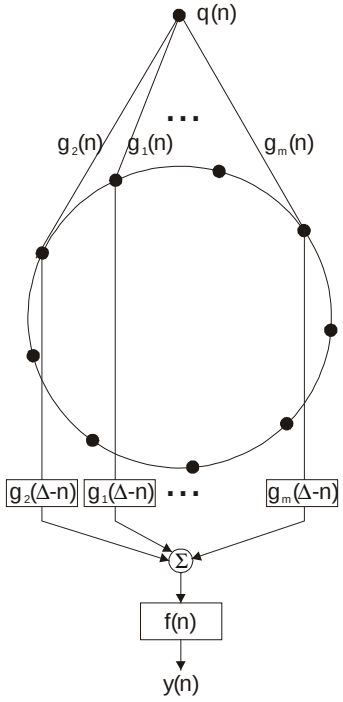

Figure 1 System used to perform the simulations

3.1 Radiation of Sound from a Point Monopole Source

The acoustic radiation from the point monopole source is given by

$$p = \frac{j\rho_0 \omega q e^{\ jkr}}{4\pi r} \tag{1}$$

where p is the complex pressure, q is the complex volume velocity (source strength), r is the distance from the source to the receivers, ρ_0 is the density, c_0 denotes the sound speed and $k=\omega/c_0$ where ω is the angular frequency. A time dependence of $e^{j\omega t}$ is assumed.

3.2 Scattering of Sound by a Rigid Sphere

When a plane wave has an obstacle in its path, in addition to the undisturbed plane wave there is a scattered wave, spreading out from the obstacle in all directions, distorting and interfering with the plane wave. The scattered wave is the difference between the actual sound wave and the undisturbed wave, which would be present if an obstacle was not there[13].

From Abramowitz *et al* [14],

$$\frac{\sin(kR)}{kR} = \sum_{0}^{\infty}(2m+1)j_m(kr)j_m(k\rho)P_m(\cos\theta) \tag{3}$$

$$\frac{\cos(kR)}{kR} = \sum_{0}^{\infty}(2m+1)j_m(kr)n_m(k\rho)P_m(\cos\theta) \tag{4}$$

where $R = \sqrt{r^2 + \rho^2 - 2r\rho\cos\theta}$, P_m is the m'th order Legendre polynomial. j_m and n_m are the m'th order spherical Bessel functions of the first and second kind respectively.

Since $exp(-j\theta)=cos\theta-jsin\theta$, it can be deduced that Equation (1) can be written as,

$$p_{ff} = -\frac{j\rho_0 k}{4\pi}\sum_{m=0}^{\infty}(2m+1)j_m(ka)(j_m(k\rho) - j\cdot n_m(k\rho))P_m(\cos\phi) \tag{5}$$

where a is the distance from the sphere centre to the receiver, ρ is the distance from the centre of the sphere to the source and ϕ is the angle subtended between the receiver point, centre of sphere and the source.

The scattered field p_s is expanded into an infinite series of waves radiating outwards, away from the origin, as proved by the $jn_m(ka)$ sign, in the following expression [15],

$$p_s = \frac{j\rho_0 k}{4\pi}\sum_{m=0}^{\infty}(2m+1)b_m(j_m(ka) - j\cdot n_m(ka))P_m(\cos\phi) \tag{6}$$

Only the outgoing waves are considered for the scattered pressure since if the converging waves were considered the radiation condition at infinity would be violated.

The constant b_m needs to be chosen in order to satisfy the boundary condition, which states that the total pressure gradient must be zero on the sphere surface. Thus,

$$b_m = \frac{j_m(k\rho) - jn_m(k\rho)}{1 - j\left(n_m'(ka)/\, j_m'(ka)\right)} \tag{7}$$

where the prime denotes differentiation with respect to the argument of the function.

It has been found that the infinite series given by Equation (6) converge satisfactorily after about m=60 terms[19]. Some examples of the total frequency response function, computed from the sum of the free field Green function and the scattered field Green function, are shown in Figure 2 for a range of angular positions on the surface of the sphere relative to the position of the point monopole source. The response was calculated at a distance of 2m between the point monopole source and the sphere centre, the sphere had a radius of 9cm. The frequency response was computed at 4096 discrete frequencies from 1Hz to 41kHz.

A Hanning window was applied to the frequency responses, in order to reduce the ringing of the impulse responses, and then the result of this computation was transformed into equivalent discrete time impulse responses, shown in Figure 3, at an effective sampling frequency of 41kHz.

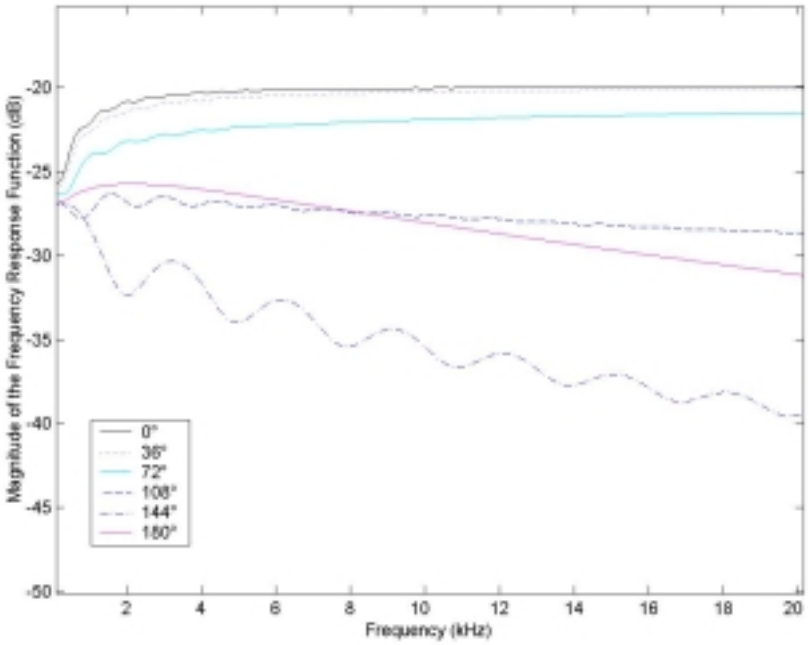

Figure 2 Total Green function in the frequency domain

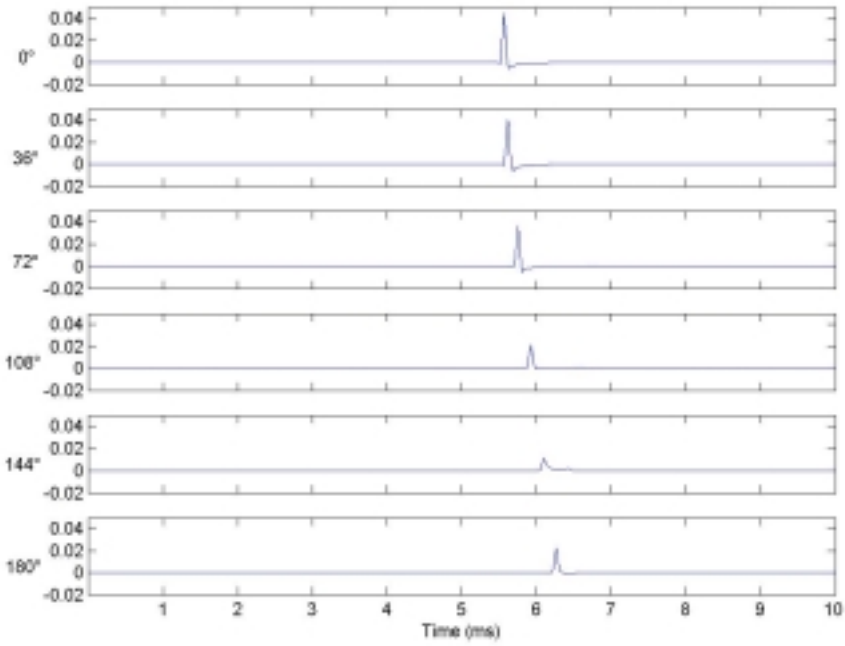

Figure 3 Total Green funtion in the time domain

3.3 Signal Processing for Accurate Replication of Source Strength Time Histories

The reception of sound in large rooms such as conference rooms and auditoria is typically contaminated by interfering noise sources and reverberation. Interference sources include

unwanted speech signals as well as acoustic noise. Beamforming methods can be used to enhance the amplitude of the desired speech signal while reducing the effects of the interfering signals and reverberation[16]. Beamforming is a space–time operation since the desired and interfering signals often originate from different spatial locations, therefore spatial separation can be exploited to separate signal from interference using a spatial filter at the receiver and if the desired signal and interferers occupy the same temporal frequency band, then temporal filtering cannot be used to separate signal from interference. Implementing a temporal filter requires processing of data collected over a temporal aperture. Similarly, implementing a spatial filter requires processing of data collected over a spatial aperture. There are two types of beamforming, namely conventional, or "fixed-weight", beamforming and adaptive beamforming. A conventional beamformer uses fixed weights and are appropriate in applications where the spatial locations of noise sources are known and not changing. In adaptive beamformers the weights are adapted to accommodate the changing signal environment[17].

3.3.1 Analytical Model

The input signal vector at the receivers is given by the sum of the signals from the source $\mathbf{s}(n)$, having a strength time history $q(n)$, and the noise $\mathbf{n}(n)$, $\mathbf{g}$ is the vector of acoustic frequency responses[12],

$$\mathbf{x}= \mathbf{s}+\mathbf{n}=Q(e^{j\omega})\mathbf{g}+\mathbf{n} \tag{8}$$

The cost function that minimises the sum of squared errors is given by[12],

$$J = E\left[\left|Y(e^{j\omega}) - Q(e^{j\omega})\right|^2\right] \tag{9}$$

where $E[\ ]$ denotes the expectation operator, $Y(e^{j\omega})=\mathbf{w}^{\mathrm{H}}\mathbf{x}$ is the output of the system in the frequency domain, and $\mathbf{w}^*$ is the vector of beamformer filters. It thus follows that

$$J = E\left[(\mathbf{w}^H\mathbf{x} - Q(e^{j\omega}))(\mathbf{w}^H x - Q(e^{j\omega}))^*\right] \tag{10}$$

Since $(\mathbf{w}^{\mathrm{H}}\mathbf{x})^*=(\mathbf{x}^{\mathrm{T}}\mathbf{w}^*)^*=\mathbf{x}^{\mathrm{H}}\mathbf{w}$ and defining $\mathbf{S}_{xx} = E\left[\mathbf{x}\mathbf{x}^{\mathrm{H}}\right]$, $\mathbf{S}_{xq} = E\left[\mathbf{x}Q^*(e^{j\omega})\right]$ and $S_{qq} = E\left[\left|Q(e^{j\omega})\right|^2\right]$, then

$$J = \mathbf{w}^{\mathrm{H}}\mathbf{S}_{xx}\mathbf{w} - \mathbf{w}^{\mathrm{H}}\mathbf{S}_{xq} - \mathbf{S}_{xq}^{\mathrm{H}}\mathbf{w} + S_{qq} \tag{11}$$

The value of the vector $\mathbf{w}$ that minimises this function is given by[12],

$$\mathbf{w}_{\mathrm{opt}} = \mathbf{S}_{xx}^{-1}\mathbf{S}_{xq} \tag{12}$$

If $\mathbf{S}_{xq}=\mathbf{g}S_{qq}$ and $\mathbf{S}_{xx}=S_{qq}\mathbf{g}\mathbf{g}^H+S_{nn}$, then

$$\mathbf{w}_{\mathrm{opt}} = \left[S_{qq}\mathbf{g}\mathbf{g}^{\mathrm{H}} + S_{nn}\right]^{-1}\mathbf{g}S_{qq} \tag{13}$$

and therefore

$$\mathbf{w}_{\mathrm{opt}} = \left[\mathbf{g}\mathbf{g}^{\mathrm{H}} + \gamma\mathbf{I}\right]^{-1}\mathbf{g} \tag{14}$$

where γ is defined as the ratio of noise spectral density, S_{nn}, to the source spectral density, S_{qq}.

This expression can also be written in an alternative form by using the matrix inversion lemma[12], given by

$$(\mathbf{A}+\mathbf{BCD})^{-1}=\mathbf{A}^{-1}-\mathbf{A}^{-1}\mathbf{B}(\mathbf{C}^{-1}+\mathbf{DA}^{-1}\mathbf{B})^{-1}\mathbf{DA}^{-1} \quad (15)$$

where the following substitutions are applied, $\mathbf{A}=\gamma\mathbf{I}$, $\mathbf{B}=\mathbf{g}$, $\mathbf{C}=1$, $\mathbf{D}=\mathbf{g}^{H}$,

$$[\mathbf{g}^{H}\mathbf{g}+\gamma\mathbf{I}]^{-1}=(1/\gamma)\mathbf{I}-(1/\gamma)^{2}\mathbf{g}(1+(1/\gamma)\mathbf{g}^{H}\mathbf{g})^{-1}\mathbf{g}^{H} \quad (16)$$

After some algebraic manipulation it follows that

$$\mathbf{w}_{opt}=\left[\gamma+\mathbf{g}^{H}\mathbf{g}\right]^{-1}\mathbf{g} \quad (17)$$

The complex conjugate of $\mathbf{w}_{opt}$ when no noise is present, $\gamma=0$, is

$$\mathbf{w}_{opt}^{*}=\left[\mathbf{g}^{H}\mathbf{g}\right]^{-1}\mathbf{g}^{*} \quad (18)$$

The solution for this optimal filter is reduced to the “delay and sum” beamformer. The “delay and sum” beamformer, also called “data independent” array signal processing method, takes the sampled sensor outputs and applies delays to these sequences in order to “best match” the incoming wave field, thereby maximising the net beamformer output when all the channels are summed together[12].

It should also be noted that the solution for the optimal beamformer is essentially that used in the focused beamformer where an attempt is made to map the distribution of acoustic source strength associated with a given source distribution by changing the assumed vector **g** of Green functions in accordance with the assumed source position[12].

The filter *f* in the system, illustrated in Figure 1, only represents the term $[\mathbf{g}^{H}\mathbf{g}]^{-1}$, since the filters $\mathbf{g}^{*}$ are used to operate on the microphone outputs. As $\mathbf{g}^{H}\mathbf{g}$ is a scalar, in Equation (18) there is only a simple division, although care must be exercised in ensuring that this filter is realisable, as discussed below.

A time reversal version of the Green function can be used to calculate $\mathbf{g}^{*}$ [18]. Thus

$$\mathbf{G}^{*}(e^{j\omega})=\mathbf{G}(e^{-j\omega})=\mathrm{FT}\{\mathbf{g}(-n)\} \quad (19)$$

The complex conjugate of the Green function equals the Fourier Transform of the time reversed Green function in the time domain. Therefore the value of $\mathbf{g}^{*}$ is calculated by taking the filter **g** impulse responses, reversing them and applying a delay to avoid problems caused by non-causal systems.

3.3.2 Results

Equation (18) was used to calculate the inverse filter *f* at 4096 discreet frequencies from 1Hz to 41kHz and the output time domain of the system was also calculated. The impulse response of the inverse filter showed a great deal of ringing in the time domain.

To improve the time domain response a regularisation parameter[20] was introduced in Equation (18).

$$\mathbf{w}_{opt}^{*}=\left[\mathbf{g}^{H}\mathbf{g}+\beta\right]^{-1}\mathbf{g}^{*} \quad (20)$$

Figure 4 shows an improvement in both the inverse filter and the output signal responses. There was a considerable decrease in the inverse filter ringing and the output signal of the system is closer to a delta function.

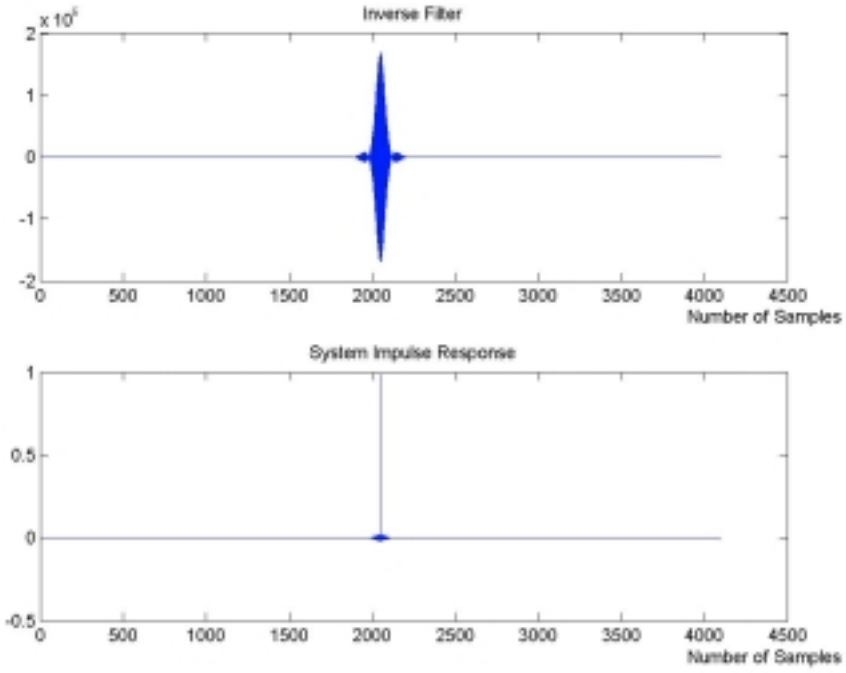

Figure 4 Inverse filter and system impulse response with regularisation

4 DIRECTIVITY

The simulations were undertaken with the source in the near field and far field. In array theory the common rule of thumb to determine the distance at which the far-field approximations begin to be valid is $R=2L^2/\lambda$, where L is the sphere diameter and λ is the acoustical wavelength[21]. Therefore 0.5m was chosen to be the arbitrary distance between the source and the centre of the arrays in the near field, and in the far field, 5m was the distance used. The aperture size of the obstacles in all three arrays are the same, therefore the width of the linear array is the same as the circumference of the circular array. The circular array radius is chosen to be 9cm, therefore the linear array will have the width of 56.55cm. When the distance within the receivers is higher than $\lambda/2$ "spatial aliasing" occurs and significant sidelobes are produced in the response of the array[12]; to plot the directivity up to 20kHz it was necessary to have 66 receivers. All the directivities were calculated with increments of 1º in the source, using 4096 samples at a sampling frequency of 41kHz.

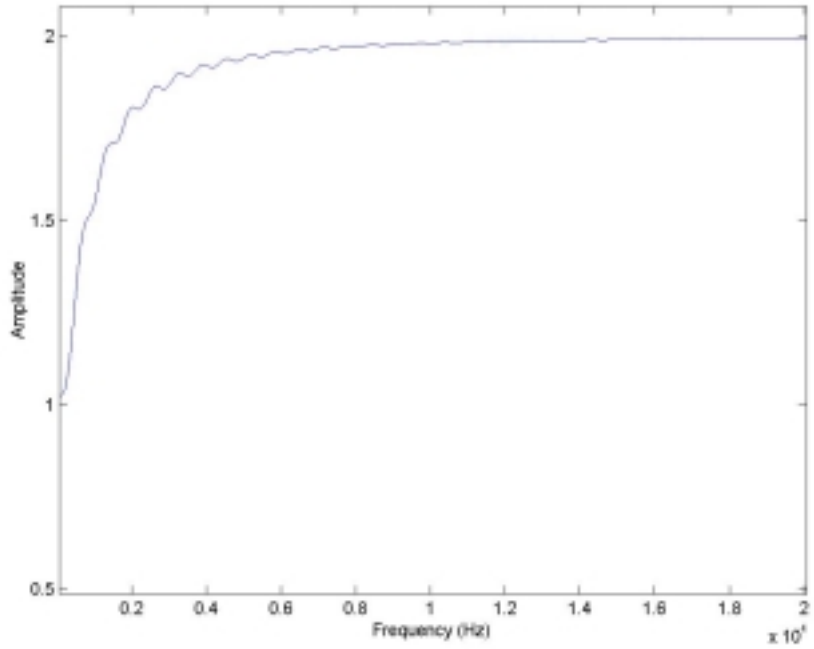

Figure 5 Total Green function over the free field Green function

When plotting the total Green function over the free field Green function, in Figure 5, it can be verified that the scattered field acts as frequency increases. Therefore it is expected that the directivity of the circular microphone array in the rigid sphere differs from the directivity of the circular microphone array in the free field as the frequency goes up.

From the directivity results shown from Figure 6 to Figure 11 it can be perceived that the linear microphone array shows rear grating lobes, which are reduced in the circular microphone arrays. When using circular arrays in the free field those rear grating lobes are reduced up to 20dB, although side lobes increase up to 15dB. By placing the circular microphone array on a rigid sphere the rear grating lobes are reduced up to 30dB while the side lobes only increase up to 10dB, therefore improving the directivity of the microphone array.

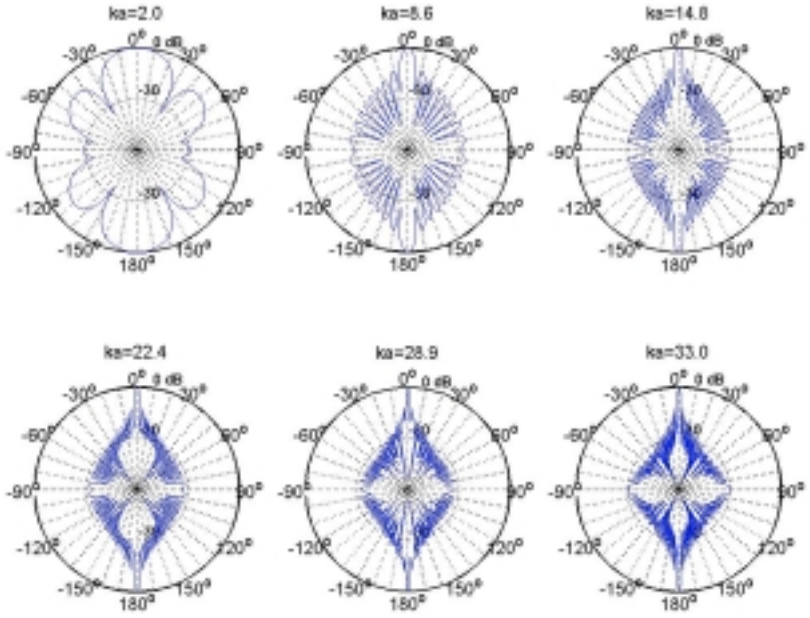

Figure 6 Directivity of a linear microphone array in the near field

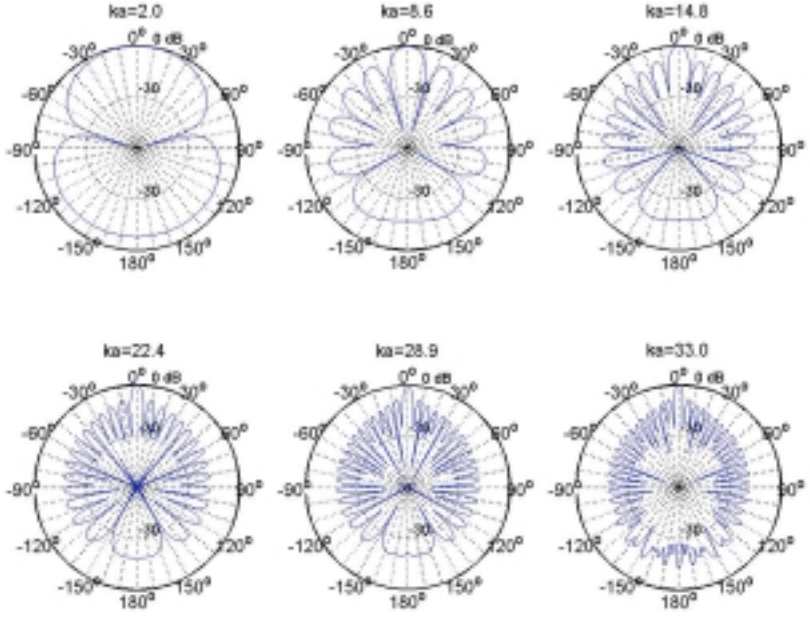

Figure 7 Directivity of a free field circular microphone array in the near field

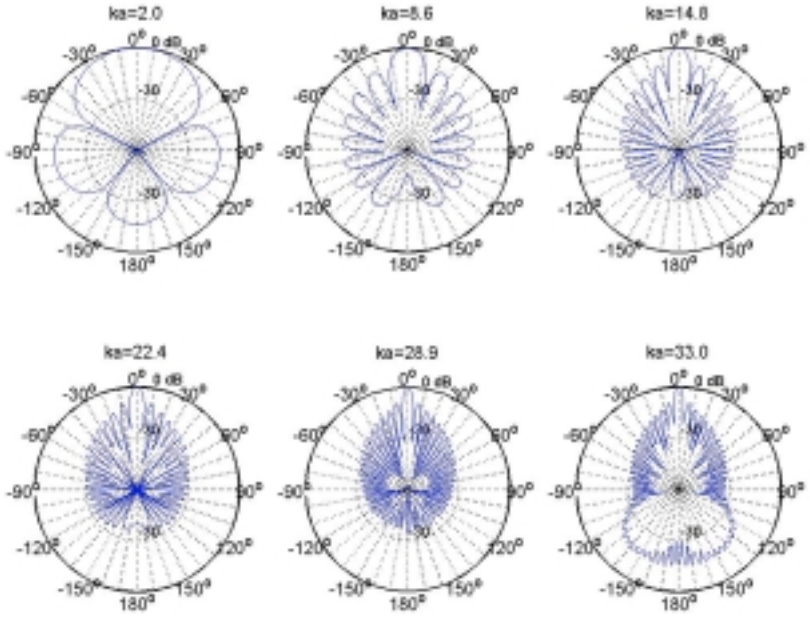

Figure 8 Directivity of a rigid sphere circular microphone array in the near field

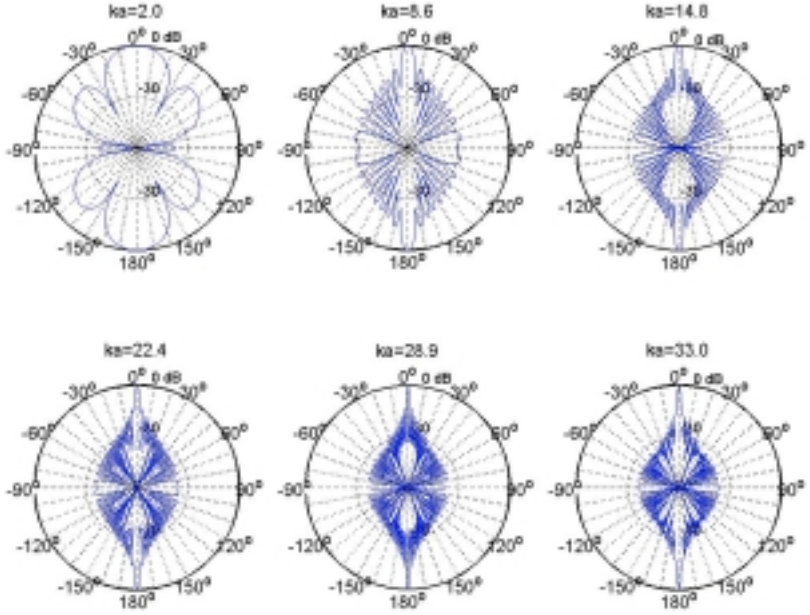

Figure 9 Directivity of a linear microphone array in the far field

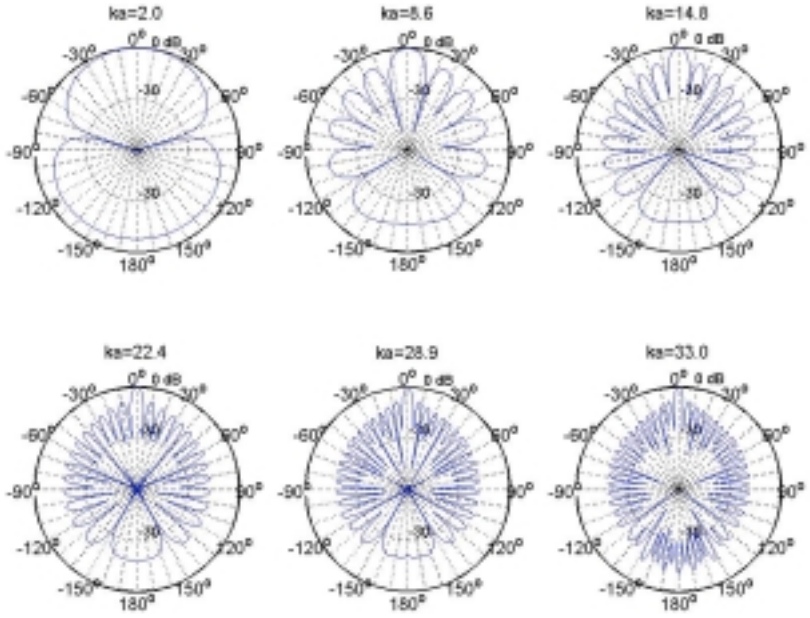

Figure 10 Directivity of a free field circular microphone array in the far field

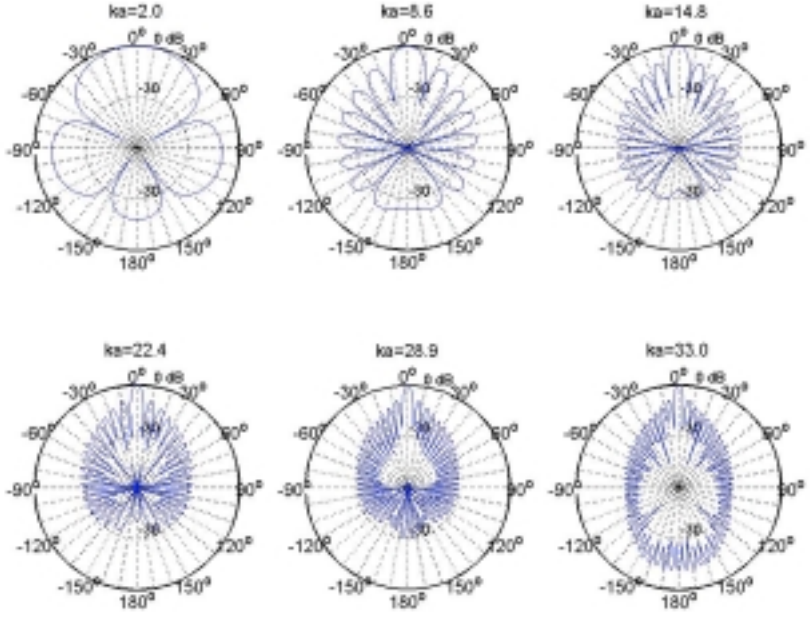

Figure 11 Directivity of a rigid sphere circular microphone array in the far field

5 CONCLUSIONS

Numerical experiments were undertaken, within a MATLAB environment, which compared the directivity performance of linear microphone arrays and circular microphone arrays. It was shown

that linear microphone arrays show a lack of ability to discriminate between front and back that is eliminated when using circular microphone arrays. The scattering of sound by a rigid sphere is shown to improve the directivity of the arrays. The beamformer used in the study was a focused beamformer that attempts to map the distribution of acoustic source strength associated with a given source distribution by changing the Green function in accordance with the assumed source position. Regularisation was added to the beamformer in order to improve the impulse responses of the filter and the output.

6 REFERENCES

1. M.J. Evans, A.I. Tew and J.A.S. Angus, 'Spatial audio teleconferencing - which way is better?', The Fourth International Conference on Auditory Display, Palo Alto, California. (November 1997).
2. Y. Mahieux, G. Le Tourneur and A. Saliou, 'A microphone array for multimedia workstations', J. Audio Eng. Soc. 44 (5). (May 1996).
3. J. Billingsley, and R. Kinns, 'The Acoustic telescope', J. Sound and Vibration 48(4) 485-510. (1976).
4. M.S. Brandstein and H.F. Silverman, 'A practical methodology for speech source localization with microphone arrays', Computer Speech and Language 11(2) 91-126. (April 1997).
5. D. Zotkin, R. Duraiswami, P. Vasanth and L.S. Davis, Smart Videoconferencing, Institute for Advanced Computer Studies, University of Maryland. (2000).
6. N. Strobel, S. Spors and R. Rabenstein, 'Joint audio-video object localization and tracking, a presentation of general methodology', IEEE Signal Processing Magazine 18(1) 22-31. (January 2001).
7. B.N. Gover, J.G. Ryan and M.R. Stinson, 'Microphone array measurement system for analysis of directional and spatial variations of sound fields, J. Acoust. Soc. America 112(5) Pt. 1. (November 2002).
8. J.M. Rigelsford and A. Tennant, 'A 64 Element Acoustic Volumetric Array', Applied Acoustics 61 469-475. (2000).
9. J. Meyer, 'Beamforming for a circular microphone array mounted on spherically shaped objects', J. Acoust. Soc. America 109(1) 185-193. (January 2001).
10. J. Meyer and G.W. Elko, 'Echoes, the newsletter of the acoustical society of America, Electroacoustic Systems for 3-D Audio – A report from the Pittsburgh meeting' 12(3). (Summer 2002).
11. J. I. Mohammad, Recording Techniques for 5-Channel Virtual Acoustic Imaging, M.S.c. Thesis University of Southampton. (May 2002).
12. P.A. Nelson, Source Strength Estimation in the Presence of Noise', unpublished notes, ISVR, University of Southampton (2002).
13. P.M. Morse, Vibration and Sound, McGraw-Hill Book Company. (1948).
14. M. Abramowitz, I.A. Stegun, Handbook of Mathematical Functions, Dover, New York. (1965).
15. O. Kirkeby and P.A. Nelson, 'Local sound field reproduction using two closely spaced loudspeakers', J. Acoust. Soc. Am. 104(4) 1973-1981. (October 1998).
16. K. Farrel, R.J. Mammone and J.L. Flanangan, 'Beamforming microphone arrays for speech enhancement', IEEE. (1992).
17. V.K. Madisetti and D.B. Williams, The Digital Signal Processing Handbook, CRC Press IEE Press. (1998).
18. A.V. Oppenheim, R.W. Shaffer, Digital Signal Processing, Prentice-Hall. (1975).
19. J.F.W. Rose, Visually adaptive virtual acoustic imaging, PhD Thesis, University of Southampton. (2003).
20. A. Tikhonov and V. Arsenin, Solution of ill-posed problems, Winston, Washington D.C. (1977).
21. C.A. Balanis, Antenna theory: analysis and design, Wiley, New York. (1997).